

University of Dundee

Real-time 3D tracking of laparoscopy training instruments for assessment and feedback

Gautier, Benjamin; Tugal, Harun; Tang, Benjie; Nabi, Ghulam; Erden, Mustafa Suphi

Published in:
Frontiers in Robotics and AI

DOI:
[10.3389/frobt.2021.751741](https://doi.org/10.3389/frobt.2021.751741)

Publication date:
2021

Licence:
CC BY

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Gautier, B., Tugal, H., Tang, B., Nabi, G., & Erden, M. S. (2021). Real-time 3D tracking of laparoscopy training instruments for assessment and feedback. *Frontiers in Robotics and AI*, 8, [751741].
<https://doi.org/10.3389/frobt.2021.751741>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Real-time 3D tracking of laparoscopy training instruments for assessment and feedback*

Benjamin Gautier, Harun Tugal, Benjie Tang, Ghulam Nabi, Mustafa Suphi Erden

Abstract— Assessment of minimally invasive surgical skills is a non-trivial task, usually requiring the presence and time of expert observers, including subjectivity and requiring special and expensive equipment and software. Although there are virtual simulators that provide self-assessment features, they are limited as the trainee loses the immediate feedback from realistic physical interaction. The physical training boxes, on the other hand, preserve the immediate physical feedback, but lack the automated self-assessment facilities. This study develops an algorithm for real-time tracking of laparoscopy instruments in the video cues of a standard physical laparoscopy training box with a single fisheye camera. The developed visual tracking algorithm recovers the 3D positions of the laparoscopic instrument tips, to which simple colored tapes (markers) are attached. With such system, the extracted instrument trajectories can be digitally processed, and automated self-assessment feedback can be provided. In this way, both the physical interaction feedback would be preserved and the need for the observance of an expert would be overcome. Real-time instrument tracking with a suitable assessment criterion would constitute a significant step towards provision of real-time (immediate) feedback to correct trainee actions and show them how the action should be performed. This study is a step towards achieving this with a low cost, automated, and widely applicable laparoscopy training and assessment system using a standard physical training box equipped with a fisheye camera.

Index Terms— Laparoscopy training, real-time motion tracking, Cartesian position estimation, single view camera, skill assessment

I. INTRODUCTION

Laparoscopy is a minimal invasive surgery performed in the abdominal cavity with the most important advantage of fast recovery of patients, compared to conventional open surgery procedures. Using only small incisions, the surgeon can perform an operation such as removing parts on organs or retrieving tissue samples for further analysis, without fully opening the abdomen [1]. However, this method brings new challenges to the surgeon as it is more difficult to perform than a conventional open surgery. The main challenges are a

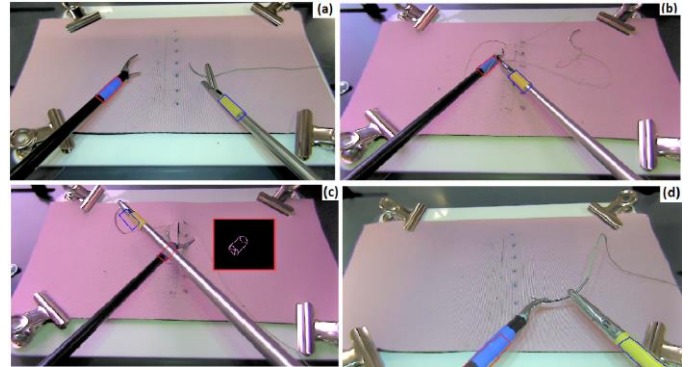


Fig. 1. (a) Instrument detection, (b) convex Hull area when the rope is crossing the blue marker, (c) convex Hull area when the instrument is crossing the blue marker and obscuring the marker, (d) out of the field of view, when Kalman filter prediction is used [27].

reduced field of view due to the use of a single camera, loss of depth perception, less sensitive force perception, and inverted motions due to a rotation around the insertion point (fulcrum effect) [2, 3]. Exemplary camera views of a suturing training from the inside of the training box used in this study are seen in Fig. 1, as adapted from our previous work [27]. To adapt to those challenges a surgeon must carry out an intensive training, which is difficult to be objectively assessed due to the lack of consistent quantitative measures [4].

Laparoscopy training aims at motor learning [5, 6] for manipulation skills with the laparoscopic instruments. Long training periods and expensive resources are required for training and evaluation of novice surgeons [7, 8, 9]. Suturing is considered to be one of the procedures that require high degree of manipulations skills in laparoscopy [10, 11]. A major concern in laparoscopy training is about evaluating the degree of skill of surgeons. Mostly offline evaluation techniques are used [12, 13], with criteria such as the number of movements of the tool coded by acceleration and deceleration thresholds, the path length covered by the tool-tip, the time taken to bring the tool-tip from one point to another [14, 15, 16], and the frequency content of time frames [17]. These criteria make use of the translational tool-tip trajectory.

For training and assessment of laparoscopy skills there are physical box trainers [12, 18, 19, 20], visual simulators [21, 22], and recently also augmented reality systems [23]. The pros and cons of these systems have been discussed in literature [23]. While box trainers provide physically realistic interaction, they require supervision by an expert for training and assessment. Virtual simulator, on the other hand, are limited in physical realism [24, 25], but allow collection of

*This research was partially funded by EPSRC under the Grant Reference EP/P013872/1.

Benjamin Gautier and Harun Tugal were, and Mustafa Suphi Erden is with Sensors, Signals and Systems, School of Engineering and Physical Sciences, Heriot-Watt University, EH14 4AL, and with Edinburgh Centre for Robotics, Edinburgh, UK. E-mail: gautier161@gmail.com; haruntugal@gmail.com; m.s.erden@hw.ac.uk

Benjie Tang and Ghulam Nabi are with Dundee Institute for Healthcare Simulation, Ninewells Hospital, and Medical School of University of Dundee, DD1 9SY, Dundee, UK. E-mail : b.tang@dundee.ac.uk ; g.nabi@dundee.ac.uk

digital data that can be processed to perform quantified assessment without the need for a supervisor. Augmented reality systems as in [23] constitute an attempt to bring together the advantages of the two: physical realism of a training box and digital computation of registered data. However, currently such systems [23] yet come with extra sensors and function with a virtual setup, though with physical instruments.

A promising approach that has been appearing in the recent years is to equip physical training boxes with machine vision and intelligence to assess the physical performance of the trainee [26, 40]. However, the emergent systems such as in [26], yet provide assessment/feedback only after the task is completed; in other words, they process the performance offline. With similar spirit, we developed in an earlier study an off-line trajectory tracking algorithm of laparoscopy tool tips in a laparoscopy training box and provided novel assessment methods using the extracted trajectories [27]. In the current paper, we present a substantially improved version of our tracking algorithm, which is capable of real-time tracking of the 3D position of instruments with a single camera, and we assess the real-time tracking performance with a Robotic Surgery Trainer setup. Our motivation is that the real-time extracted instrument trajectories can be digitally processed, and automated self-assessment feedback can be provided to the trainee in real-time. In this way both the physical interaction feedback would be preserved and the need for the observance of an expert would be overcome by provision of instant feedback when the trainee makes a mistake or deviates from the optimal way of performing the task.

In this study, we apply our algorithm to the videos of training sessions for intra-corporeal wound suturing, which is considered to be one of the most difficult procedures in laparoscopy training [4, 28, 29]. For that purpose, we have recorded videos from six professional surgeons and ten novice subjects. Ethical approval was acquired from the Ethics Committee of School of Engineering and Physical Sciences at Heriot-Watt University with Ethics Approval number 18/EA/MSE/1 and all participants provided their Informed Consent prior to data collection. Our real-time tracking algorithm in this paper is successful to extract the same trajectories from these recorded videos as the off-line algorithm we presented in our previous work [27]. Therefore, the conventional assessment criteria we used in [27] from the literature [4, 30, 31, 32, 33] and the novel one we proposed in [27] are all applicable also to the real-time extracted trajectories in the current study, highlighting the usefulness of the trajectories to distinguish between novice and professional performances. We do not repeat the explanation and application of these criteria in this paper and refer the reader to [27].

Tracking methods for objects in known environments are well known in the literature and have already been used in several studies on laparoscopy, such as 2D tooltip location tracking in laparoscopy training videos for eye-hand coordination analysis [34], 3D laparoscopy instrument detection using the vanishing point of the edges of the instrument's image [35], stereo-imaging with two webcams and markers [36], monochrome image processing [37], making use of the position of the insertion points of the

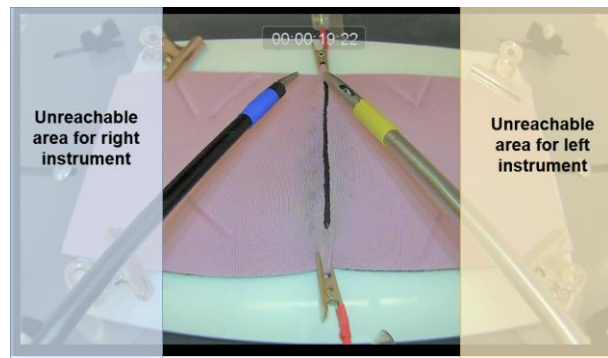


Fig. 2. Practical workspace for both instruments based on the collected data.

instruments [38, 39], optical flow information in video frames [40], and instrument tracking for calibration purposes for robotic surgery [41]. Among these, the ones that target training mostly use colored markers on the tips of the instruments to be tracked. This is justified for training setups as it is easily applicable to any training laparoscopy instrument and it does not impact the performance of the subject. However, a major challenge with marker-based instrument tracking is that the markers might be obscured or they might get out of the field of camera view [42], as illustrated in Fig. 1. In this paper we also develop a marker based tracking system; but in comparison to the other methods, *i)* we address the problem of occlusion and disappearance from the scene by adopting a Kalman filter to estimate the position only in such instances of disappearance from the scene, and *ii)* we do the tracking for 3D positioning of two instrument tips in real-time with a speed of 25 frames per second by using the geometric features of the markers. We achieve real-time tracking purely based on a single camera image processing from a standard laparoscopy training box. As our system does not add any extra equipment to a standard laparoscopy training box, we consider it to be low-cost and widely applicable as it can easily be applied to any training box.

The rest of the paper is organized as follows. Our implementation of the real-time tracking in 2D images using color-based markers is presented in Section II. In Section III, we explain the method used for 3D Cartesian position estimation in real-time. In Section IV, we compare the performance of trajectory extraction with respect to the ground truth trajectories generated by a Robotic Surgery Trainer setup incorporating two UR3 Universal Robots. Section V concludes the paper.

II. MARKER CORNER DETECTION IN 2D IMAGES

Instrument detection is realized by tracking the colored tapes attached to the end of the two instruments as in Fig.1 and Fig. 2. The colors of the tapes are chosen to be easily separable from the background (usually a pink colored suturing pad) and each other in a Hue Saturation Value (HSV) space. The tracking problem in this setting can be described as subject to a close-to-invariant light exposure (closed environment and short time recording). The image processing techniques used in this study are individually well known in literature; therefore, we will only mention them briefly without detailed explanations. We note that, what we have performed in this

study is adapting these techniques and integrating them effectively to solve the specific detection challenges in the context of a laparoscopy training practice. For example, and specifically, the method we have developed allows detection of the corners of a markers even when some parts of the marker are separated from each other in the image, which can happen in two different cases: when the rope is wrapped around the instrument over the tracked marker as in Fig. 1(b), and when one of the instruments obscures part of the tracked marker on the other instrument as in Fig. 1(c). The real-time tracking process is realized using a two steps method, a *detection* step where the four corners of a colored tape on each instrument are found in the current 2D image frame (explained in the following sub-sections) and a *tracking* step where a 3D position of the instrument tips are generated using the detected corners (explained in Section III).

A. Preprocessing

Using the recorded video retrieved from our experiments, an HSV color database is constructed for the range of the pink background pad and for the range of the colors of the tapes on the instruments, across 20 videos recorded from 6 professional and 10 novice subjects. The HSV range for the tapes is identified by isolating 50x50 square regions containing each tape. The mean HSV values are then extracted and used to create the database to be compared to new inputs [43, 44]. The current database comprises three different illumination setting across the 20 videos: a natural light recording setup, and two artificial light recording environment, one in our laboratory and one in the medical facilities. When a new pad is used in any lighting condition, its detected color is compared to the dataset using a minimum distance formula, and the closest corresponding HSV range is selected for each instrument for the detection.

The second part of the preprocessing is a full frame detection. Using color space conversion (cvtColor in OpenCV) on the full image is time consuming; therefore, it is performed only on the first two video frames. The marker positions in the image are retrieved and the center of gravity of each detected contour around the marker is used to estimate the position of the contour in the next frame using the motion gradient. Finally, the extrinsic parameters of the camera are retrieved using the perspective transformation matrix based on the pink pad background corners in the very first frame. Then the Euler angles representing the camera orientation are extracted, as the camera height being adjustable and hence might change across the use of the system in different times. In our setup this initialization is applied automatically every time the system is turned on.

B. Detection of Corners in 2D Image

Our detection process follows a general framework for HSV object detection [45, 46] with a real-time adjustable detection window for time efficiency. The embedded camera has a 25 frames-per-second (fps) reading rate thus the full process needs to be designed to have a minimum of 50-Hz response to generate the position of two instrument tips in each frame cycle. In order to obtain the Cartesian information at a rate of

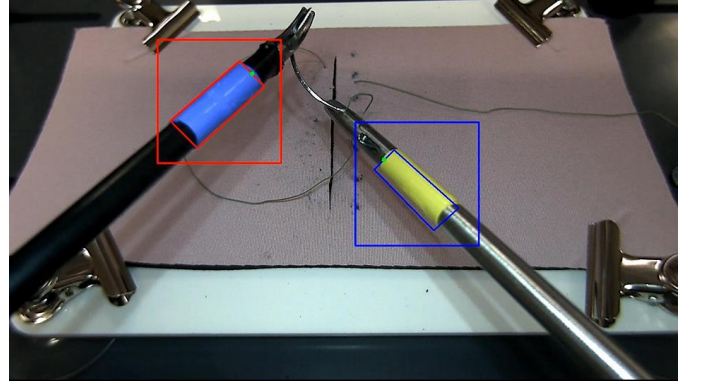


Fig. 3. Sample ROI squares around the markers.

minimum 50 Hz, the detection process is only applied on a windowed section of the frame centered around the estimated position of the instruments. The center of gravity of the marker tapes (\mathbf{c}_g) in the next image frame is estimated using the gradient of motion found in the region of interest (ROI) in the previous frames. ROI is defined as a square window where we estimate the marker to be inside in each frame (Fig. 3). Using the estimate of the gradient of motion in the previous frame, a new ROI is generated in each frame and the search for the tip position is performed only in this ROI, rather than the whole image. The size of the single edge of the ROI window, ω , which is adjusted in each step, is computed as follows:

$$\omega = \left[a_{prev} \times \left(\frac{\omega_{max} - \omega_{min}}{a_{max}} \right) + \omega_{min} \right] \times \left[\frac{\|\nabla \mathbf{c}_g\|}{c_{g,vmax}} + 1 \right] \quad (1)$$

where a_{prev} is the area of the detected marker in the previous image, a_{max} is the empirically identified maximum area of a marker in an image when the marker is closest to the camera (18,000 pixels), ω_{max} is the maximum edge size of ROI set to 400 pixels, ω_{min} is the minimum edge size of ROI set to 100 pixels, and $c_{g,vmax}$ is an empirically chosen value for the maximum speed of the tool tip across frames set to 30 pixels per frame. The factor on the left hand side of the equation handles the size of the window of ROI based on the prior knowledge of the instrument size and the factor on the right hand side adjusts the size of the ROI based on the velocity of the instrument computed on the previous frames. The estimation of the \mathbf{c}_g and adjustment of the size of the search window (ROI) for each frame are the keys to achieving a fast-enough detection allowing real-time tracking.

In order to apply a region of interest with the dynamic size detection process, some specific cases must be defined and handled properly to avoid wrong detection and thus losing the instruments. The specific cases are identified as follows:

- The area is not consistent with the previous detection.
- The velocity of the instrument is not realistic.
- The detection algorithm could not find a set of four corners in the previous ROI.

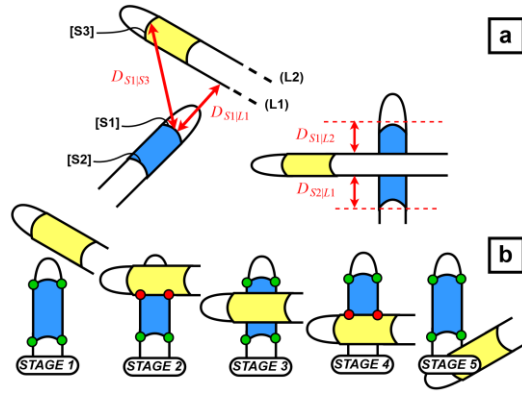


Fig. 4. (a) Distance between instruments; (b) crossing between instruments.

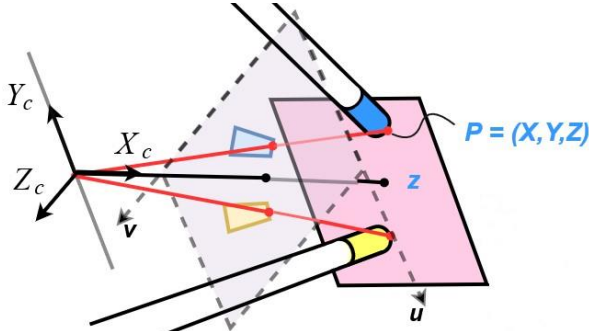


Fig. 5. Representation of the trapezoidal shapes after flattening the view of the cylinders and the angle of the instruments regarding the camera.

If any of these special cases is detected, the next position estimation is rendered incorrect and the ROI cannot be computed, thus a search of the instrument is applied on a larger part of the image. This detection also is not made on the full image. A notion of dead space is identified based on the recorded dataset which leads to identification of a “practical workspace” for the instruments as in Fig. 2 and the search is conducted only in this workspace.

One of the challenges with two instrument detection is the crossing event when one of the instruments is obscured or partially obscured by the other (Fig. 1(c), Fig. 4). To deal with such crossing, the separated parts of the overall contour of the obscured marker are detected using Canny detection and then a convex hull is created using the detected parts. This method allows a regrouping of the separated parts to deal with the situation illustrated in Stage 3 of the crossing (Fig. 4(b)). Furthermore, it allows to simplify the representation of the geometry of the contour and thus speeds up the process of finding the corners using Hough transform.

Another advantage of using a convex hull representation is the following. In order to speed up the detection process, we are directly using the raw (non-flattened) fisheye camera output, thus the shape retrieved before application of the convex hull does not have straight lines. This would result in having a large set of candidate points for corner selection after the Hough transform. This is avoided and the number of the candidate points for the corners is narrowed down by adapting a convex hull. Using the raw feed also results that we cannot directly identify the correct set of 4 points in the Hough transform output. To overcome this, we flatten the output

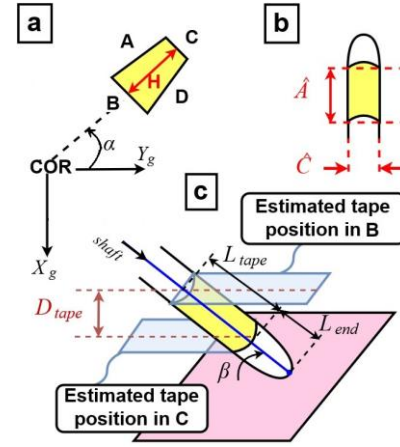


Fig. 6. (a) Estimated α angle, rotation around Z of the instrument around the Center of Rotation and parameters of the trapezoidal detection; (b) parameters of the real instruments; (c) estimated β angle, rotation around X.

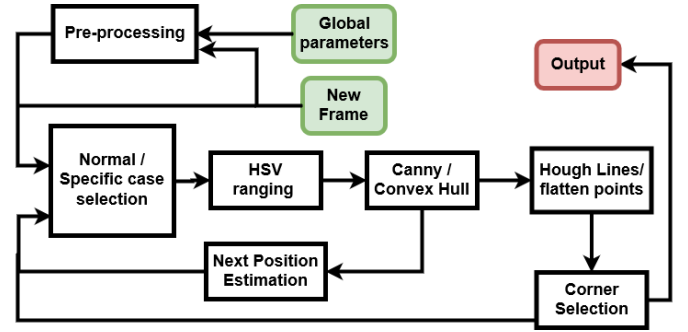


Fig. 7. Diagram of the detection process with image processing.

points from the Hough transform. In this way, we apply flattening only to about 40 points at the end of the detection, instead of approximately 2 million initially in the frame. For flattening, we use the intrinsic parameters of the camera (distortion, focal lengths and focal points) retrieved in advance from a chessboard calibration.

For the corner selection we use the knowledge on the contour pose in the image and the properties of the trapezoidal shape when the cylinders (markers) are viewed from top (Fig. 5) where the detected line segments C and B in Fig. 6(a) must remain parallel. This method allows to find the best candidates for the corner points from the list output in the previous steps. The full detection process with image processing can be summarized with the block diagram in Fig. 7.

C. Estimation for Missing Corners

In the previously mentioned specific cases, a correct corner detection is not possible with the image processing as explained up to this point. In these cases, a Kalman filter is used to estimate the position of the instruments. The Kalman filter implementation is a standard one where the estimation of the next position of the instrument is based on a corrector and predictor equation [47, 48]. The corrector uses the previous measurement to update the model and the predictor estimates the next position using the error covariance of the model. In our application, we use the Kalman filter estimation only when marker corners cannot be found through previously explained image processing procedures. Deciding on the fly

when to switch between the actual detection and Kalman filter prediction is not a trivial task. Usually the algorithms that detect occlusions use the geometric properties of the object and therefore are computationally expensive. In this study a different approach is developed using the specificities of the laparoscopy training environment and specifically the distance between the two instruments. The markers can be missing in the laparoscopy training context in two cases, first, some corners being occluded by the crossing of the instruments (Fig. 4(b)) and second, the instrument being outside the field of view (Fig. 2(d)). For the first case, our method relies on the detection of the two instruments being close enough to each other for one of them being occluded by the other. For that purpose, we compute the minimal distance between a detected segment and the line formed by the other segment or the tip of the other instrument, using the following formulas:

$$\text{sign}(d_{s_j|L_i}) = \text{sign}\left(\frac{a_i \cdot x_{m_j} - y_{m_j} + b_i}{\sqrt{1+a_i^2}}\right) \quad (2)$$

$$\text{sign}(d_{s_j|s_i}) = \text{sign}\left(\sqrt{(x_{m_j} - x_{m_i})^2 + (y_{m_j} - y_{m_i})^2}\right) \quad (3)$$

where a_i and b_i are the parameters of the line $L_{i=1,\dots,4}$ and (x_{m_j}, y_{m_j}) is the middle point of a given segment s_j as described in the Fig. 4(a), and d stands for distance. The sign of the computed distances between the edge lines in the above formulas can be used to detect the occlusions. When an occlusion is detected as such, the Kalman filter estimation is used. Furthermore, in order to reduce the use of the Kalman filter prediction, it is possible to detect the instances when recovery of the actual corners of the markers is possible during the occlusion of the marker body. In Fig. 4(b), we can see that during Stage 3, the corners of the markers can be completely recovered.

The second situation that requires the Kalman filter estimation is when one of the instruments goes out of the video frame which can simply be estimated and detected using the minimal distance between the previously detected location of the instrument and the borders of the image frame. For such cases, the Kalman filter estimation is directly used. Finally, the four corners of both instruments estimated by the Kalman Filter are flattened similarly as in the previous sub-section.

D. Overall Procedure of Detection of Marker Corners

The following are the overall steps of procedure applied for detection of the marker corners in the 2D images as explained in this section:

1. Check whether any of the “specific cases” applies (such as the detection could not find four corners in the previous ROI); if so, perform Pre-processing in the practical workspace (Fig. 2); otherwise continue with Step 2.
2. Identify the new ROI (Equation (1)).
3. Apply HSV decomposition in the ROI.
4. Apply Canny edge detection
5. Apply Convex Hull regrouping to obtain a connected contour representing the marker.
6. Apply Hough transform for line detection.
7. Identify the candidate corners and apply flattening at these corner points.
8. Identify the four corners of the marker using the information of geometric relations.
9. Check if any marker is occluded (Equation (2) and (3)) or out of view; if, yes, use Kalman Filter output to estimate the location of the occluded corners and apply flattening at the estimated corners; otherwise stay with the identified corners in Step 9.
10. Output the corner coordinates for depth estimation.

III. TOOL TIP POSITION TRACKING IN 3D

In order to track the full 3D position of the tip of the instruments, we first reconstruct the 2D position information from the 2D Camera view following the methods explained in Section II. Afterwards, we estimate the depth using the difference between the computed circumference of the detected marker polygon as seen in the image and the actual circumference of the polygon.

A. Real-time 3D Tracking

As seen in Fig. 2, there are two instrument tips, each with four degrees of freedom actively controlled by the subject. However, in this study we track only the three degrees of freedom, the translational movements of each instrument, and ignore the rotational movement around the shaft axis. This is because, almost all criteria of performance that apply to instrument movements in laparoscopy training [12]-[17] make use of the 3D position of the instrument tips, but not the orientation of the tip. The tip point trajectories without the orientation provide a rich enough information for assessment purposes in laparoscopy training exercises.

The three degrees of freedom translational movement of the tip point can be represented by (or translated into) other movement parameters, possibly some of them defined as rotations around specific axes, such as rotation of the instrument shaft around an axis through the insertion point. In this study, we consider successive elementary transformations with respect to the “current reference frame” constructed after each transformation [49]: specifically, a rotation of the instrument shaft at the insertion point with an angle α around the z axis of the ground frame, rotation with an angle β around the x axis of the intermediary frame, and a translation of the tip point along the instrument shaft in y axis of the successive intermediary frame, as shown in Fig. 6. These three motion parameters can easily be translated into the tip point translation parameters along the three Cartesian axes of a global reference frame through straightforward geometric relations. Let \mathbf{R}_0 be the orthogonal global reference frame with x and y axes parallel to the ground and its origin at the instrument center of rotation (COR) (the insertion point) and \mathbf{R}_F be the reference frame located at the tip of the instrument ($\mathbf{R}_{F_{i-1}}$ and \mathbf{R}_{F_i} for each instrument, respectively), as illustrated in Fig. 8. A homogeneous transformation matrix, ${}^0\mathbf{T}_F$, can be computed in between the \mathbf{R}_0 and \mathbf{R}_F reference frames in terms

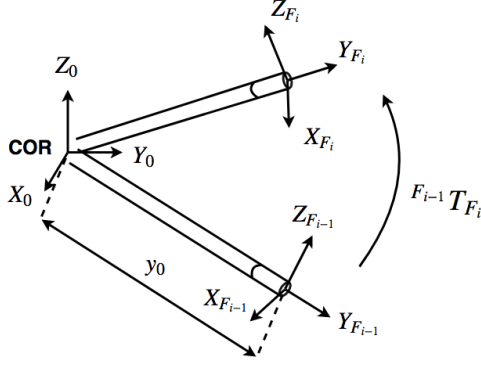


Fig. 8. Reference frames at the center of rotation and tool tip [27].

of the mentioned elementary rotation and translation matrices as,

$${}^0\mathbf{T}_F = \text{Rot}(z, \alpha) \cdot \text{Rot}(x, \beta) \cdot \text{Trans}(y, y_0) \quad (4)$$

to transform the representation of a point in \mathbf{R}_0 to that in \mathbf{R}_F as,

$$\begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ 1 \end{pmatrix} = {}^0\mathbf{T}_F \begin{pmatrix} x_F \\ y_F \\ z_F \\ 1 \end{pmatrix}. \quad (5)$$

Tracking the tip point of an instrument corresponds to identifying the α , β , and y_0 parameters in the above transformation matrix, which can then be used to find the Cartesian position of the tip of the instrument with respect to the global reference frame \mathbf{R}_0 . Our approach is first to identify the α and β angles, using the computed depth difference between the front (C edge in Fig. 6(a)) and rear (B edge in Fig. 6(a)) segments of the marker along with the geometric relations as shown in Fig. 6. The depth estimation, d , is based on the ratio of the perimeter of the detected marker in the 2D image to the actual perimeter of the marker as in Equation (6),

$$d = \frac{p_{Tool|real}}{p_{Tool|img}} \quad (6)$$

where $p_{Tool|real}$ represents the actual perimeter of the marker and $p_{Tool|img}$ represents the computed perimeter from the image. First, we consider the front segment, the edge C in Fig. 6(a) and use this as the image correspondence of the actual length \hat{C} of the marker. We already know the ratio between \hat{A} and \hat{C} segments of the actual marker in Fig. 6(b) and using this ratio we can compute a length A_C as the image correspondence of the actual edge, \hat{A}_i . Using these, the perimeter of the rectangle in the image at the location of C can be computed as $2(C + A_C)$. We can then compute the depth of the rectangle parallel to the ground as located at C (Fig. 6(c)). Following the same procedure, we can also compute the depth of the rectangle parallel to the ground as located at B . The difference between these two depth values provide us with the depth difference due to inclination, D_{tape} value, indicated in Fig. 6(c). Knowing the actual width, \hat{A} and the depth

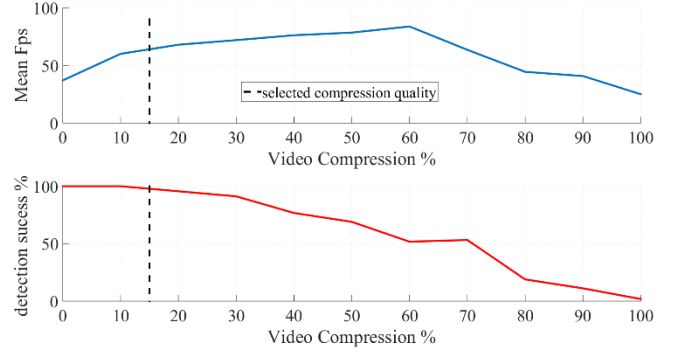


Fig. 9. The average speed of image processing to detect the corners of a marker on a single instrument in terms of frames-per-second (upper figure) and the rate of correct detection (lower figure) with respect to varying compression rate of video frames transmitted from the camera to the computer.

difference we can compute the β angle. For α , the angle between the line connecting the centers of C and B edges and the y axis of the global reference frame is computed (Fig. 6(a)). Once β is known, the actual length of the instrument can be computed considering the visible length in the image and the inclination angle β , along with the ratio between the actual length and the visible length in the image when the instrument is straightly aligned parallel to the ground (perpendicular to the camera view).

B. Testing for Real-Time Processing

In this section we present our analysis of the speed of processing of the overall algorithm in terms of frames-per-second (FPS), with respect to the compression rate we use in streaming the video to the computer and considering the success rate of detection of the corner points of the marker at an instrument-tip.

In order to achieve a fast processing, we use a streaming communication (TCP/IP) between the camera and the computer housing the image processing software. For that purpose, we apply a compression process on the video feed (on the slave side) to ensure fast and smooth streaming prior to tracking (on the master side). The rate of compression for the streaming is a major factor that impacts the overall speed and performance of detection. We use a JPEG compression and Fig. 9 presents the results depicting the speed and performance of detection with varying compression rate. As it is observed in this figure, below 60% compression, the speed of processing increases whereas the performance for correct detection decreases monotonically. In this graph, 15% compression seems to be an optimal choice to achieve a sufficiently fast speed (above 50 Hz) and a high rate of correct detection (very close to 100%); therefore, we applied 15% compression throughout the tests presented in the following section.

IV. TESTING AND VERIFICATION

In our previous study [27] we had tested our off-line tracking algorithm with human subject experiments, where a subject manually manipulated one of the instruments to make its tip to follow the edges of a rectangular object with known dimensions, and where we used the shape and dimensions of

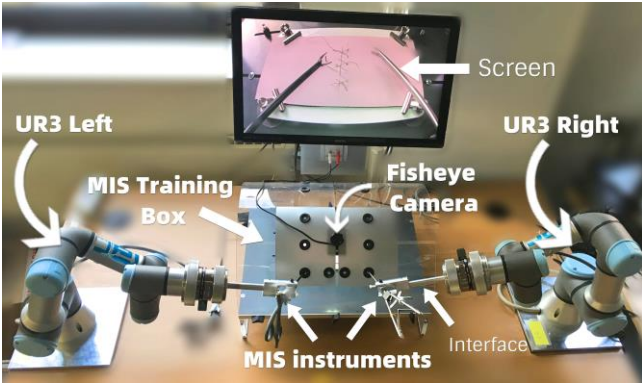


Fig. 10. Robotic Surgery Trainer setup used in this study to test the position accuracy of the real-time tracking algorithm.

the object as a reference for measurement. That method did not distinguish between the actual measurement error of the image processing algorithm and the deviation of the trajectory from the edges of the box due to human hand tremor. Therefore, in the current study we make the measurements with a robotic manipulation setup, the Robotic Surgery Trainer system in our lab incorporating two UR3 Universal Robots to manipulate the laparoscopy instruments (Fig. 10). With this setup we can accurately record the ground truth positions of the tip of the instruments through the position data provided by the encoders of the robots.

In order to compare the instrument tip trajectory recorded by the robot to that estimated by the real-time tracking algorithm, we first transform the trajectory retrieved from the video tracking into the robot base frame. We then synchronize the two datasets as the robot recording frequency is 125Hz, giving us a larger number of points in the robot trajectory compared to the tracked trajectory on the video. We eliminate the Euclidian distance between the numeric values of robot recorded and tracked trajectories considering the initial and final points of the trajectories, in order to align them as closely as possible. We then apply a zero-padding in frequency domain to equate the sample size of the position data in the two trajectories. Finally, we apply a norm distance measure between the data of every corresponding couple in the two trajectory data sets to find out the maximal distance between the two trajectories.

For this measurement, we again considered the boxes we had used in the previous study [27] (Fig. 11 (a)), but this time instead of tracing the actual edges of the physical boxes, we made the robots generate the motions to follow the edges of hypothetical boxes with the tip of the instrument, without the physical presence of the box. In this way all six edges were reachable by the instrument as in Fig. 11 (b). In this figure the red lines show the trajectory followed by the tip of the instrument as recorded by the robot and blue lines show the estimated trajectory as tracked in real-time by the image processing algorithm presented in this paper. We used four different rectangular boxes: a small box occupying half of the screen, a thin box occupying half of the screen, a large box occupying a large space in the screen, and a large and thin box occupying a large space in the screen. Those experiments were realized using both left-hand and right-hand instruments.

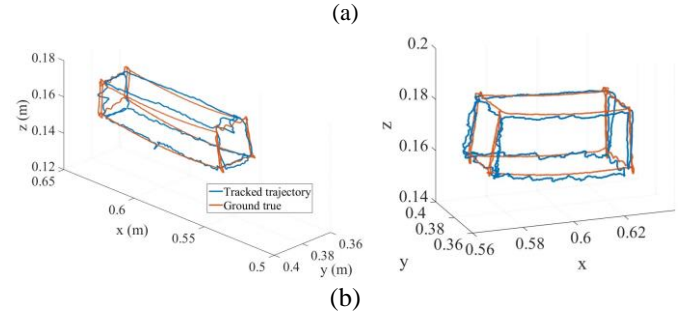
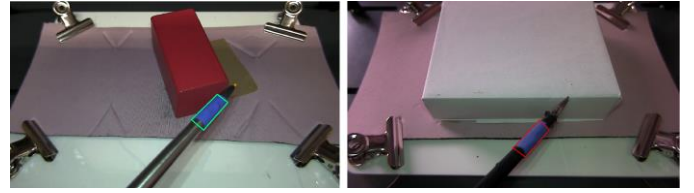


Fig. 11. (a) Two sample boxes used to generate the trajectories. (b) The trajectories generated by the robot (red) and tracked by the real-time image processing algorithm (blue) (units: m).

The maximum error between the estimated trajectories compared to the robot recorded trajectories through all the experiments was computed to be 1.5 mm in x , 4 mm in y , and 3 mm in z coordinates along the edges of the boxes as in Fig. 11 (b). This performance is sufficient for our purposes to assess skill level with typical criteria as we applied in [27].

V. CONCLUSION

In this paper a real time 3D instrument trajectory tracking is developed for single camera laparoscopy training boxes. Trajectories extracted in real-time would be useful to perform real-time skill assessment and to provide real-time feedback, immediately as the subject performs unskilled motions. The work here is a first step towards achieving that goal, as it provides the facility for real-time trajectory extraction. The next step to build on this work would be to develop the assessment criteria that would function in real-time and that would be in such a characteristic to provide immediate feedback to the trainee. The criteria that would serve that purpose are yet to be developed and tested. In our previous work [27], we demonstrated a novel criterion based on the detection of the spatial distribution of the tip positions of the right-hand and left-hand instruments, which functioned significantly superior to existing conventional criteria in literature to distinguish between professional and novice performances we had recorded. The criterion is mainly based on spatial positions of the tips and checks whether the right-hand instrument tip is in its required region in the right-hand side section in the box, and does the same for the left-hand instrument (Fig. 12). As this criterion does not rely on history of the positions, we consider it to be promising to be adapted with the presented real-time tracking algorithm to instantly check the performance and generate useful real-time feed-back to the trainee. The real-time tracking algorithm developed in the present study and a potential adaptation of the assessment criterion presented in [27], or similar others yet to be developed, together would be a significant step towards a self-training system with real-time feedback, which would

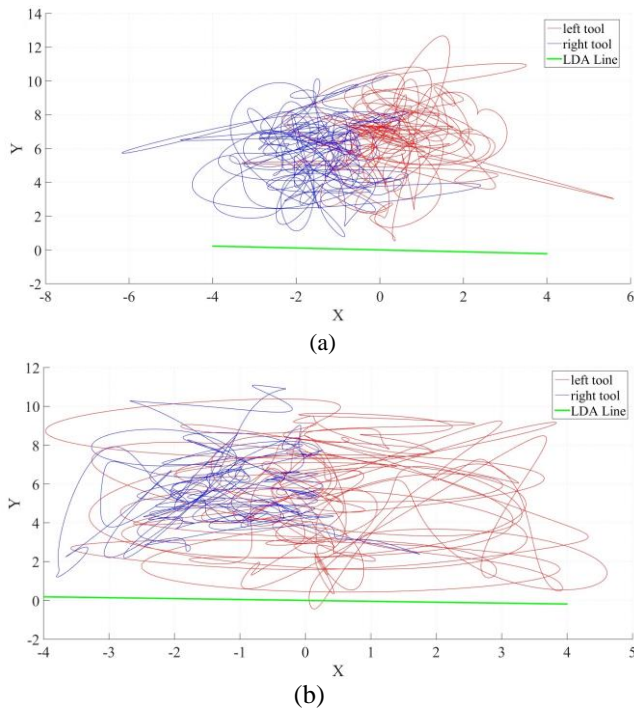


Fig. 12. Sample laparoscopy instrument tip-point trajectories (units: cm) as successfully discriminated to belong (a) to a professional and (b) to a novice by our novel assessment criterion based on Linear Discriminant Analysis (LDA) presented in [27]. The right-hand instrument (the driver) trajectory is in red and the left-hand instrument (the receiver) trajectory is in blue. The LDA line in green shows the best direction to distinguish the right-hand and left-hand instruments according to their spatial distribution and as expected it is almost the same in each case in this specific suturing exercise; reflecting that the orientation of the suturing line is the same and perpendicular to the axis that separates right and left hand tools.

eliminate the need for an expert human trainer, would be low-cost, and would be widely applicable with standard and single camera laparoscopy training boxes. Our future work will progress in this direction.

REFERENCES

- [1] D. Eyvazzadeh and S. M. Kavic, "Defining Laparoscopy Through Review of Technical Details in JSLS," *Journal of the Society of Laparoscopic & Robotic Surgeons*, vol. 12, pp. 151–153, 2011.
- [2] H. Xina, J. S. Zeleka, and H. Carnahanb, "Laparoscopic surgery, perceptual limitations and force : A review," CSCBC, 2006.
- [3] B. Levy, "Principles of safe laparoscopic surgery," *Surgery*, vol. 35, no. 4, pp. 216–219, 2017.
- [4] O. H. Chang, L. P. King, A. M. Modest, and H.-C. Hur, "Developing an objective structured assessment of technical skills for laparoscopic suturing and intracorporeal knot tying," *Journal of Surgical Education*, vol. 73, no. 2, pp. 258 – 263, 2016.
- [5] Halsband, U. and Lange, R.K, "Motor learning in man: a review of functional and clinical studies", *Journal of Physiology – Paris*, 99: 414-424, 2006.
- [6] Dankelman, J., Chmarra, M.K., Verdaasdong, E.G.G., Stassesn, L.P.S., and Grimbergen, C.A., "Fundamental aspects of learning minimally invasive surgical skills", *Minimally Invasive Therapy*, 14 (4-5): 247-256, 2005.
- [7] Gutt, C.N., Kim, Z.G., and Krahenbuhl, L., "Training for advanced laparoscopic surgery". *Eur. J. Surg*, 168: 172-177, 2002.
- [8] Vassiliou, M.C., Feldman, L.S., Andrew, C.G., Bergman, S., Leffondre, K., Stanbridge, D., and Fried, G.M., "A global assessment tool for evaluation of intraoperative laparoscopic skills". *The American Journal of Surgery*, 190 (1): 107-113, 2005.
- [9] Berg, D.A., Milner, R.E., Fisher, C.A., Goldberg, A.J., Dempsey, D.T., and Grewal, H., "A cost-effective approach to establishing a surgical skills laboratory". *Surgery*, 142: 712-721, 2007.
- [10] Judkins, T.N, Oleynikov, D., Narazaki, K., and Stergiou, N, "Surgical Endoscopy", 20 (5): 824-829, 2006.
- [11] Wang, H., Wang, S., Ding, J., and Luo, H., "Suturing and tying knots assisted by a surgical robot system in laryngeal mis", *Robotica*, 28 (2): 241-252, 2010.
- [12] Aggarwal, R., Moorthy, K., and Darzi, A., "Laparoscopic skills training and assessment", *British Journal of Surgery*, vol. 91 (12), pp.1549-1558, 2004.
- [13] Chmarra, M.K., Grimbergen, C.A., and Dankelman, J., "Systems for tracking minimally invasive surgical instruments", *Minimally Invasive Therapy*, 16 (6): 328-340, 2007.
- [14] Datta, V., Mackay, S., Mandalia, M., and Darzi, A., "The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model", *J. American College of Surgeons*, 193 (5): 479-485, 2001.
- [15] Moorthy, K, Munz, Y., Dosis, A., Bello, F. and Darzi, A., "Motion analysis in the training and assessment of minimally invasive surgery", *Minimally Invasive Therapy and Allied Technologies*, 12 (3-4): 137-142, 2003.
- [16] Chmarra, M.K, Jansen, F.W., Grimbergen, C.A., and Dankelman, J., "Retracting and seeking movements during laparoscopic goal-oriented movements. Is the shortest path length optimal?", *Surgical Endoscopy*, 22: 943-949, 2008.
- [17] Megali, G., Sinigaglia, S., Tonet, O., and Dario, P., "Modelling and Evaluation of Surgical Performance Using Hidden Markov Models", *IEEE Trans. on Biomedical Eng.*, 53 (10): 1911-1919, 2006.
- [18] Schreuder, H., van den Berg, C., Hazebroek, E., Verheijen, R., Schijven, M., "Laparoscopic skills training using inexpensive box trainers: which exercises to choose when constructing a validated training course," *BJOG-An International Journal af Obstetrics and Gynaecology*, vol.118, pp.1576-1584, 2011.
- [19] Kunert, W., Storz, P., Dietz, N., Axt, S., Falch, C., Kirschniak, A., and Wilhelm, P., "Learning curves, potential and speed in training of laparoscopic skills: a randomised comparative study in a box trainer," *Surgical Endoscopy And Other Interventional Techniques*, 2020, DOI: 10.1007/s00464-020-07768-1.
- [20] Ulrich, A.P., Cho, M.Y., Lam, C., Lerner, V.T., "A Low-Cost Platform for Laparoscopic Simulation Training", *Obstetrics and Gynecology*, vol. 136 (1), pp. 77-82, 2020.
- [21] Hedman, L., Nisell, H., Fellander-Tsai, Li., Enochsson, L., Ahlborg, L., and Felländer-Tsai, L., "Simulator training and non-technical factors improve laparoscopic performance among OBGYN trainees," *Acta obstetricia et gynecologica Scandinavica*. vol. 92(10), p.1194-1201, 2013.
- [22] Strandbygaard, J., Bjerrum, F., Maagaard, M., Winkel, P., Larsen, C.R., Ringsted, C., Gluud, C., Grantcharov, T., Ottesen, B., and Sorensen, J.L., "Instructor Feedback Versus No Instructor Feedback on Performance in a Laparoscopic Virtual Reality Simulator A Randomized Trial", *Annals of Surgery*, vol. 257 (5), pp. 839-844, 2013.
- [23] Lahanas, V., Loukas, C., Smailis, N., and Georgiou, E., "A novel augmented reality simulator for skills assessment in

- minimal invasive surgery”, *Surgical Endoscopy*, vol. 29, pp. 2224-2234, 2015.
- [24] von Websky, M.W., Raptis, D.A., Vitz, M., Rosenthal, R., Clavien, P.A., and Hahnloser, D., “Access to a Simulator is Not Enough: The Benefits of Virtual Reality Training Based on Peer-Group-Derived Benchmarks – A Randomized Controlled Trial, *World Journal of Surgery*, vol 37 (11), pp. 2534-2541, 2013.
- [25] Greco, E.F., Regehr, G., and Okrainec, A., “Identifying and Classifying Problem Areas in Laparoscopic Skills Acquisition: Can Simulators Help?”, *Academic Medicine*, vol. 85 (10), pp. S5-S8, 2010.
- [26] Alonso-Silverio, G.A., Pérez-Escamirosa, F., Bruno-Sanchez, R., Ortiz-Simon, J.L., Muñoz-Guerrero, R., Minor-Martinez, A., and Alarcón-Paredes, A., “Development of a Laparoscopic Box Trainer Based on Open Source Hardware and Artificial Intelligence for Objective Assessment of Surgical Psychomotor Skills”, *Surgical Innovation*, vol. 25 (4), pp. 380-388, 2018.
- [27] Gautier, B., Tugal, H., Tang, B., Nabi, G., Erden, M.S., “Laparoscopy instrument tracking for single view camera and skill assessment”, *International Conference on Robotics and Automation (ICRA)*, Palais des congres de Montreal, Montreal, Canada, May 20-24, 2019.
- [28] L. T. De Paolis, F. Ricciardi, and F. Giuliani, “Development of a serious game for laparoscopic suture training,” in *Augmented and Virtual Reality*, L. T. De Paolis and A. Mongelli, Eds. Cham: Springer International Publishing, 2014, pp. 90–102.
- [29] Hudgens, J.L. and Pasic, R.P., *Fundamentals of Geometric Laparoscopy and Suturing*, Endo-Press GmbH, Germany, 2015.
- [30] S. Estrada, C. Duran, D. Schulz, J. Bismuth, M. D. Byrne, and M. K. OMalley, “Smoothness of surgical tool tip motion correlates to skill in endovascular tasks,” *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 5, pp. 647–659, Oct 2016.
- [31] S. G. Kroeze, E. K. Mayer, S. Chopra, R. Aggarwal, A. Darzi, and A. Patel, “Assessment of laparoscopic suturing skills of urology residents: A pan-european study,” *European Urology*, vol. 56, no. 5, pp. 865 – 873, 2009.
- [32] S. N. Z. Ahmmad, E. S. L. Ming, Y. C. Fai, and F. K. bin Che Harun, “Assessment methods for surgical skill,” *World Academy of Science, Engineering and Technology International Journal of Biomedical and Biological Engineering*, vol. Vol:5, No:10, 2011.
- [33] G. Retrosi, T. Cundy, M. Haddad, and S. Clarke, “Motion analysisbased skills training and assessment in pediatric laparoscopy: Construct, concurrent, and content validity for the eosim simulator,” *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 25, no. 11, pp. 944–950, 2015.
- [34] Jiang, X., Zhend, B., and Atkins M.S., “Video Processing to Locate the Tooltip Position in Surgical Eye–Hand Coordination Tasks”, *Surgical Innovation*, vol. 22(3), pp. 285-293, 2014.
- [35] Allen, B.F., Kasper, F., Nataneli, G., Dutson, E., and Faloutsos, P., “Visual Tracking of Laparoscopic Instruments in Standard Training Environments”, *Studies in Health Technology and Informatics*, vol. 163 (18), pp. 11-17, 2011.
- [36] Pérez-Escamirosa, F., Oropesa, I., Sánchez-González, P., Tapia-Jurado, J., Ruiz-Lizarraga, J., and Minor-Martínez, A., “Orthogonal cameras system for tracking of laparoscopic instruments in training environments”, *Cirugia y Cirujanos*, vol. 86, pp. 548-555, 2018.
- [37] Zhang, X. and Payandeh, S., “Application of visual tracking for robot assisted laparoscopic surgery,” *J. Field Robotics*, vol. 19, pp. 315–328, 2002.
- [38] Doignon, C., Nageotte, F., and de Mathelin, M., “The Role of Insertion Points in the Detection and Positioning of Instruments in Laparoscopy for Robotic Tasks”, R. Larsen, M. Nielsen, and J. Sporring (Eds.): *MICCAI 2006*, LNCS 4190, pp. 527–534, 2006.
- [39] Voros, S., Orvain, E., Cinquin, P., and Long, J.A., “Automatic detection of instruments in laparoscopic images: a first step towards high level command of robotized endoscopic holders”, *IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, Pisa, Italy, 2006.
- [40] Sánchez-Margallo, J.A., Sánchez-Margallo, F.M., Pagador, J.B., Gómez, E.J., Sánchez-González, P., Usón, J., Moreno, J., “Video-based assistance system for training in minimally invasive surgery”, *Minimally Invasive Therapy & Allied Technologies*, vol. 20 (4), pp. 197-205, 2011.
- [41] Zhang, F., Wang, Z., Chen, W., He, K., Wang, Y., and Liu, Y.H., “Hand-Eye Calibration of Surgical Instrument for Robotic Surgery Using Interactive Manipulation”, *IEEE Robotics and Automation Letters*, vol. 5 (2), pp. 1540-1547, 2020.
- [42] Lin, B., Sun, Y., Qian, X., Goldgof, D., Gitlin, R., and You, Y., “Video-based 3D reconstruction, laparoscope localization and deformation recovery for abdominal minimally invasive surgery: a survey”, *Int J Med Robotics Comput Assist Surg*, vol. 12, pp. 158–178, 2016.
- [43] Riaz, M. Pankoo, K., and Jongan, P., “Extracting Color Using Adaptive Segmentation for Image Retrieval”, *Proc. of 2nd International Joint Conference on Computational Sciences and Optimization (CSO)*, Sanya, China, 2009.
- [44] Abter, S.O. and Abdullah, N.A.Z., “Colored Image Retrieval based on Most used Colors”, *International Journal of Advanced Computer Science and Applications*, vol. 8 (9): 264-269, 2017.
- [45] Cucchiara, R., Grana, C., Piccardi, M., Prati, A., and Sirotti, S., “Improving shadow suppression in moving object detection with HSV color information”, *Proc. of IEEE Intelligent Transportation Systems Conference (ITSC'01)*, pp. 334-339, 2001.
- [46] Hamuda, E., Mc Ginley, B., Glavin, M., and Jones, E., “Automatic crop detection under field conditions using the HSV colour space and morphological operations”, *Computers And Electronics In Agriculture*, vol. 133: 97-107, 2017.
- [47] Chen, S.Y., “Kalman Filter for Robot Vision: A Survey”, *IEEE Transactions On Industrial Electronics*, vol. 59 (11): 4409-4420, 2012.
- [48] Nguyen, H.T., Smeulders, A.W.M., “Fast occluded object tracking by a robust appearance filter”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26 (8): 1099-1104, 2004.
- [49] Craig, J., *Introduction to Robotics – Mechanics and Control*, Pearson Prentice Hall, 3rd Edition, pp. 19-61.